



## OPEN **Belief-consistent information is most shared despite being the least surprising**

Jacob T. Goebel<sup>1</sup>✉, Mark W. Susmann<sup>1,2,3</sup>, Srinivasan Parthasarathy<sup>3</sup>, Hesham El Gamal<sup>4</sup>, R. Kelly Garrett<sup>5</sup> & Duane T. Wegener<sup>1</sup>

In the classical information theoretic framework, information “value” is proportional to how novel/surprising the information is. Recent work building on such notions claimed that false news spreads faster than truth online because false news is more novel and therefore surprising. However, another determinant of surprise, semantic meaning (e.g., information’s consistency or inconsistency with prior beliefs), should also influence value and sharing. Examining sharing behavior on Twitter, we observed separate relations of novelty and belief consistency with sharing. Though surprise could not be assessed in those studies, belief consistency should relate to less surprise, suggesting the relevance of semantic meaning beyond novelty. In two controlled experiments, belief-consistent (vs. belief-inconsistent) information was shared more despite consistent information being the least surprising. Manipulated novelty did not predict sharing or surprise. Thus, classical information theoretic predictions regarding perceived value and sharing would benefit from considering semantic meaning in contexts where people hold pre-existing beliefs.

**Keywords** Novelty, Surprise, Belief consistency, Sharing

People encounter countless pieces of information throughout their daily lives, and online interpersonal communication plays a major role in this exposure. This presents a critical challenge; given limited time and cognitive resources, what information is deemed ‘valuable’ enough to share? Information theory suggests that information derives its value from how novel or surprising it is perceived to be<sup>1,2</sup>. Furthermore, the prevailing information theoretic approach is semantic-free, meaning that it considers only whether the information is known or not; the theory does not take into account how the information semantically fits with existing beliefs. Implicit in this perspective is the assumption that novelty and surprise are consistently positively correlated.

We argue, however, that the attributes of novelty and surprise depend on the subjective meaning of the information, specifically its consistency with prior beliefs. We refer to this as a context-sensitive perspective. Novelty is best defined as the degree to which information is not yet stored in memory, whereas surprise is an emotional response to information that contrasts with one’s expectation formed through prior experience<sup>3,4</sup>. From this perspective, novel information is not necessarily surprising. For example, people are rarely surprised by new evidence that their existing beliefs are correct. Because novelty and surprise can operate independently, they can, under some conditions, exhibit opposing relations with sharing behavior. Distinguishing novelty and surprise as antecedents of information diffusion can bolster efforts to predict large-scale social movements<sup>5–7</sup>, implement outreach campaigns and PSAs<sup>8</sup>, and combat the prevalence of online misinformation<sup>9,10</sup>, among other applications<sup>3</sup>.

There is ample evidence that information novelty can promote sharing. Antismoking messages containing novel arguments have been found to exhibit greater retransmission<sup>11</sup>, and novel health and political information regarding COVID-19 spread faster than familiar information<sup>12</sup>. Scholars have even speculated that the reason false information posted on Twitter tends to be shared more widely than true information is that those falsehoods tended to be more novel and more likely to elicit surprise<sup>13</sup>. Other research has linked surprise with information sharing<sup>14</sup>. News articles that evoke surprise go viral more often<sup>15</sup>, and potentially surprising information is often shared with limited concern for its accuracy<sup>16</sup>.

<sup>1</sup>Department of Psychology, Ohio State University, Columbus, OH, USA. <sup>2</sup>Present address: Department of Psychology, Vanderbilt University, Nashville, TN, USA. <sup>3</sup>Department of Computer Science and Engineering, Ohio State University, Columbus, OH, USA. <sup>4</sup>Faculty of Engineering, University of Sydney, Sydney, NSW, Australia. <sup>5</sup>School of Communication, Ohio State University, Columbus, OH, USA. ✉email: goebel.83@osu.edu

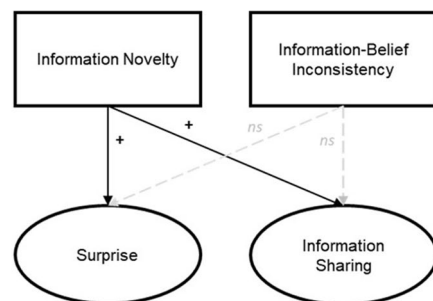
Although novelty and surprise are often positively correlated, they are distinct concepts such that surprise can be derived from sources other than novelty<sup>3,17</sup>. For instance, stories with expectation-violating twists are more surprising<sup>18</sup>, and persuasion research demonstrates that people are more likely to find information surprising when it contradicts rather than validates their pre-existing attitudes<sup>19,20</sup>. Classical information theory obscures this difference by asserting that surprise evoked by information (termed “surprisal”) is inversely proportional to its a priori probability of occurrence<sup>3,21</sup>. When dealing with human experience, this probability would be the probability of previously encountering the information. However, Bayesian theory is more nuanced. It defines surprise as a response to the *discrepancy* between prior beliefs and beliefs implied by the information<sup>22–24</sup>. Thus, a piece of information might be objectively “novel” (i.e., never seen previously) but elicit little surprise due to its consonance with prior beliefs. Similarly, information that is inconsistent with one’s beliefs might be perceived as surprising even if it is not novel<sup>3</sup>. In line with this idea, a significant body of research treats belief consistency as separable from perceived novelty<sup>25–27</sup>.

Despite their distinct antecedents, novelty and surprise have rarely been examined independently in relation to information sharing, potentially due to the prevailing notion from semantic-free information theory that connects novelty and surprise. In some cases, surprise has been used as a proxy for measuring novelty<sup>13,28,29</sup>. However, surprise, stemming from inconsistency of information with beliefs might sometimes influence information value in ways that contrast the predicted effects of novelty. Whereas novelty has been associated with greater sharing<sup>13</sup>, information running counter to strong beliefs (i.e., surprising information) is often viewed with skepticism<sup>30</sup> and shared less than information consistent with one’s beliefs or attitudes<sup>31–37</sup>. Yet, it seems likely that belief-inconsistent information would generally be viewed as more surprising. Classic information theory has largely ignored such possibilities<sup>1,2</sup>. Past research has not directly tested whether the effects of objective novelty and belief consistency on surprise (and sharing) are separable<sup>37</sup>.

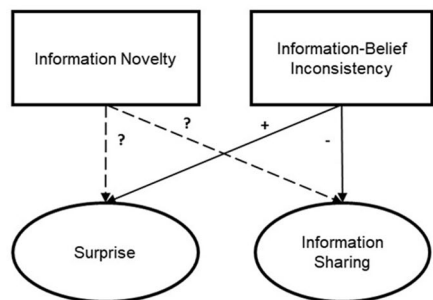
Factors other than novelty can clearly guide information sharing. For instance, despite observing that novel information was more likely than non-novel information to spread via email, some research found that non-novel content was shared more than novel content via social media<sup>29</sup>. This again suggests that factors other than novelty, perhaps including belief consistency, might exert greater influence on sharing intentions, at least in some contexts. Most Americans say that accuracy should be a very important part of their sharing decisions<sup>37</sup>. Yet the illusory-truth effect tells us that familiarity—the opposite of novelty—promotes the perception that information is accurate<sup>36,38,39</sup> and can promote sharing of fake news<sup>14,40,41</sup>.

In sum, we pit predictions based on the semantic-free perspective against those derived from the context-sensitive perspective (see Fig. 1). We argue that the effects of novelty and surprise on information sharing are context-sensitive and need not always work in parallel. Surprise is not solely based on information novelty; instead, we argue that it is also influenced by semantic consistency with prior beliefs. The first and most important aim of the present research is thus to confirm that belief consistency is a reliable predictor of information sharing and that it has an effect that is independent of information novelty. To this end, we observed information sharing behavior on Twitter, assessing the extent to which tweets’ novelty and belief consistency, two potential antecedents

### Semantic-Free Information Theory Prediction



### Context-Sensitive Prediction



**Figure 1.** Comparison between semantic-free and context-sensitive perspectives on the relations connecting information novelty and information-belief inconsistency to perceptions of surprise and information sharing.

of surprise, drive sharing. In addition, to assess the causal mechanisms behind these relations, we conducted two experiments testing (a) whether information's match (or mismatch) with one's prior beliefs induces surprise (independent of information novelty) and (b) whether belief consistency and novelty might differentially impact information sharing and perceived value.

## Results

### Twitter analyses

We conducted several analyses to assess the independent influence of information novelty and information-belief consistency on information sharing via Twitter.

#### *Novelty associated with sharing*

To assess whether novel tweets are shared more often than less novel tweets, we first identified tweets that should differ in novelty but not belief consistency. Operationalizing novelty with Twitter data is challenging; Twitter does not report which tweets a user has seen. Given this constraint, we opted to operationalize novelty by capitalizing on the fact that the first information shared about a new event will necessarily be more novel than information about that same event shared later. Although there is no guarantee that users will have seen any one tweet, earlier tweets are, in aggregate, more novel than subsequent tweets about the same event. Extending this idea, we operationalized novelty using the time since the initial tweet about a discrete event—tweets occurring soon after the event should be quite novel, whereas tweets occurring longer after the event should be less novel.

We collected politically relevant tweets from politically neutral media sources that contained clearly pro-liberal or pro-conservative information about discrete events. To test whether novel tweets (that is, those posted closer to the time of the event) were shared more frequently than less novel tweets, we constructed a general linear mixed-effects model predicting the number of retweets each tweet received using the number of minutes following the first tweet about the event to which a given tweet pertained. We also included the source of a tweet and the event to which it pertained as random factors to account for variability across sources and events. Finally, because the number of retweets is a count variable whose distribution is likely to be Poisson in form, we used Poisson modeling to estimate the effects of tweet timing on the number of retweets.

Results were consistent with novelty being associated with information sharing. The more time that had passed since the first tweet about an event was posted, the fewer retweets the information received,  $b = -3.44$ ,  $se = 0.17$ ,  $Z = -20.602$ ,  $p < 0.001$ , 95% Confidence Interval (CI)  $[-3.776, -3.121]$ .

#### *Ideological consistency associated with sharing*

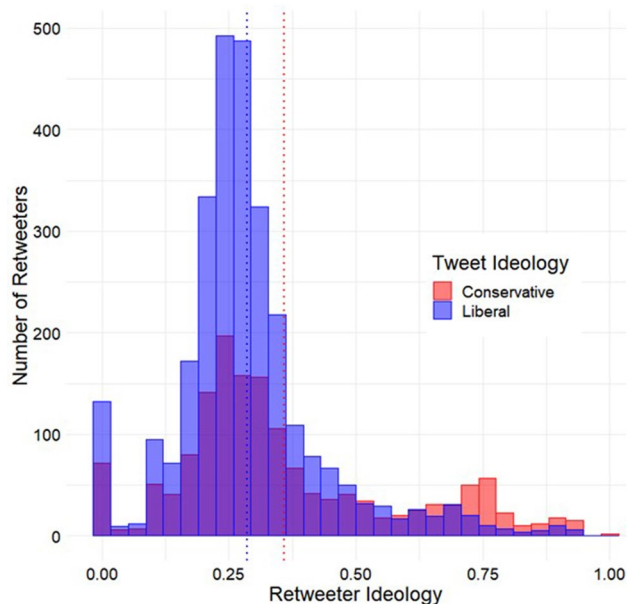
To examine whether tweets containing belief-consistent information are more likely to be shared than those containing belief-inconsistent information, we examined whether individuals were more likely to share tweets expressing views that aligned with their political ideology. We manually coded the slant of each tweet, and we estimated users' ideology as the average ideology of the U.S. politicians they followed on Twitter (see Methods for more detail)<sup>42</sup>. To facilitate the intensive analysis required for each retweeter, a subsample of 43 Tweets was drawn from the original set of 223 Tweets used in the novelty analysis. Because the ideology estimation method relied on observing which U.S. politicians a user followed, only users following at least one politician were included in these analyses. This resulted in a final sample of 4413 retweeters. Consistent with expectations, an independent samples t-test confirmed that users who retweeted liberal-aligned tweets were, on average, significantly more liberal ( $M = 0.29$ ,  $SD = 0.15$ ) than users who retweeted conservative-aligned tweets ( $M = 0.36$ ,  $SD = 0.22$ ; see Fig. 2),  $t(4411) = -13.10$ ,  $p < 0.001$ , Mean Difference 95% CI  $[-0.083, -0.062]$ .

To ensure the robustness of these results, we replicated these analyses using a second dataset by collecting an additional, non-overlapping sample of 64 political tweets that provided clearly pro-liberal or pro-conservative content (again from politically neutral media sources and about discrete events). This dataset included 5724 retweeters. Results were consistent with what we observed in the first analysis. Users who retweeted liberal-aligned tweets were significantly more liberal ( $M = 0.26$ ,  $SD = 0.13$ ) than users who retweeted conservative-aligned tweets ( $M = 0.38$ ,  $SD = 0.24$ ; see Fig. 3),  $t(5722) = -25.42$ ,  $p < 0.001$ , Mean Difference 95% CI  $[-0.138, -0.119]$ .

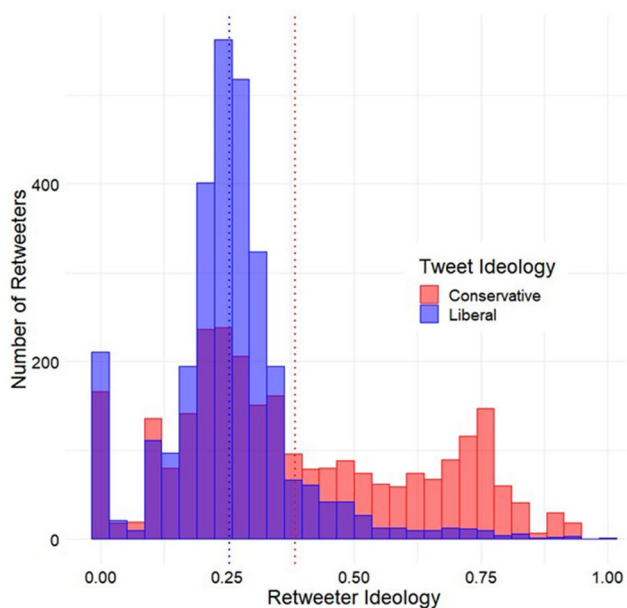
Taken together, these analyses support the notion that for a given tweet, both its novelty and its ideological consistency—which is related to belief consistency<sup>43</sup>—are plausible predictors of information sharing. Most importantly, more widespread sharing of tweets aligned rather than misaligned with one's beliefs might suggest that information value does not solely derive from novelty/surprise of the information. Rather, consistency of the information with existing beliefs might enhance the perceived value of the information despite the likelihood that such information would not be perceived as particularly surprising.

These at-scale analyses have several limitations that make it difficult to draw strong conclusions about the roles of novelty and belief consistency in information sharing. First, because Twitter does not make available information about which tweets a given individual has seen (unless they interact with it in some way), it is impossible to determine novelty at the individual level. Relatedly, like previous research using Twitter data, we do not have access to Twitter's internal algorithm for surfacing tweets to users. Thus, we do not know to what extent the patterns observed here are the product of individuals' reactions to the novelty of a tweet versus the way the Twitter algorithm incorporates novelty. Although the freshness of information seems likely to correspond with its novelty for those who encounter it, our proxy for novelty is far from ideal. Thus, it is not possible with these data to directly compare the influences of novelty and belief consistency on sharing.

Second, ideological consistency is an imperfect proxy for beliefs. Liberals and conservatives often hold different beliefs thanks to ideological perceptual biases<sup>43</sup>, but ideology is a noisy proxy for belief. For example, conservatives might be less likely than liberals to believe that human activity is contributing to climate change, but a majority of conservatives do believe that human activity contributes<sup>44</sup>. Additionally, and quite importantly



**Figure 2.** Distributions of retweeter ideology estimates as a function of tweet ideology, sampling tweets from the novelty analysis. The blue dotted line represents the mean ideology of retweeters of liberal-aligned tweets, whereas the red dotted line represents that mean ideology for retweeters of conservative-aligned tweets.



**Figure 3.** Distributions of retweeter ideology estimates as a function of tweet ideology from a second, independent sample of tweets. The blue dotted line represents the mean ideology for retweeters of liberal-aligned tweets, whereas the red dotted line represents that mean ideology for retweeters of conservative-aligned tweets.

for our purposes, there are no good measures of surprise using Twitter data. Thus, we cannot tell whether belief-consistent information was, in fact, less surprising than belief-inconsistent information or whether (lack of) surprise would relate to information sharing. Finally, sharing decisions about political events are likely driven by more than just relevant prior beliefs, including self-presentational concerns<sup>45,46</sup>, defensiveness<sup>47</sup>, and social identity or the desire to bond with others<sup>46</sup>. Examining sharing decisions in a more controlled situation where one has no ulterior motives would therefore be desirable. Much prior research explores how broad individual

ideological alignment, partisanship, or core values impact decisions to transmit information<sup>48</sup>, including most work on selective sharing<sup>32–34,36,37</sup>.

Thus, we conducted a pair of experiments that manipulated initial beliefs in order to provide more direct and causal information about belief consistency predicting information sharing. A goal of the experiments was to more cleanly instantiate participants' initial beliefs to better isolate the effect of belief consistency per se on sharing intentions. The two experiments also allowed us to directly manipulate information novelty in addition to belief consistency, as well as measure subjective novelty and surprise.

## Experiments

We conducted a pair of experiments designed to address the limitations of the observational Twitter study. In both experiments, we induced a set of beliefs. We manipulated both the novelty and the belief consistency of a message that we presented to participants, and we asked participants to assess the message and to decide whether to share it. The experiments used different topics as a form of stimulus sampling, and Experiment 2 included additional outcome measures.

We adapted a paradigm from prior research examining the perseverance of newly established beliefs<sup>49,50</sup>. We induced initial beliefs by presenting participants with information about an unfamiliar topic. In Experiment 1, the information presented (fictional) evidence that risk takers make either better or worse firefighters. In Experiment 2, the information presented (fictional) evidence that a country should or should not be allowed to join the European Union (EU). We employed topics about which people did not already have strong pre-existing beliefs in order to facilitate our manipulations of belief consistency and novelty. This enables us to more clearly examine the causal impact of beliefs unconfounded with prior experience and knowledge.

Next, we asked participants to play the part of an editorial assistant at a news organization by reviewing two new pieces of information. Each piece of information either supported or contradicted their initial beliefs (the belief consistency manipulation) and either had or had not been seen before (the novelty manipulation). Participants rated how novel and surprising both pieces of information were and indicated whether the information should be transmitted to the newspaper editor. The second experiment also included a measure of perceived information value.

### *Experiment analysis strategy*

Most analyses used multiple regression, with exceptions noted below. Save for the manipulation check, predictors in these analyses were always the same: outcomes were predicted by the consistency condition (i.e., whether the information provided in the update supported or contradicted the conclusion established in the earlier phase), as well as the novelty condition (i.e., whether information contained in the update had been previously encountered or not).

### *Manipulation check*

Our manipulations were effective in instilling the desired initial beliefs. In Experiment 1, those in the riskiness-is-good-for-firefighting condition indicated that they thought riskiness was significantly better for firefighting than those in the riskiness-is-bad-for-firefighting condition,  $b = 22.62$ ,  $se = 1.55$ ,  $t(224) = 14.58$ ,  $p < 0.001$ , 95% CI [19.566, 25.682],  $r = 0.70$ . In Experiment 2, those provided with favorable country attributes indicated that they thought the country should be allowed to join the EU significantly more than those provided with less favorable country attributes,  $b = 18.60$ ,  $se = 1.21$ ,  $t(299) = 15.34$ ,  $p < 0.001$ , 95% CI [16.211, 20.982],  $r = 0.66$ .

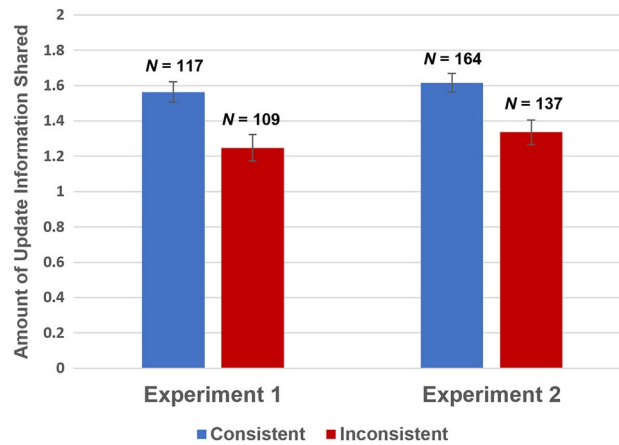
### *Belief inconsistency associated with surprise*

Belief-consistent information was rated as less surprising than belief-inconsistent information in both Experiment 1,  $b = -0.61$ ,  $se = 0.08$ ,  $t(223) = -7.39$ ,  $p < 0.001$ , 95% CI [-0.766, -0.444],  $r = 0.44$ , and Experiment 2,  $b = -0.70$ ,  $se = 0.10$ ,  $t(298) = -7.22$ ,  $p < 0.001$ , 95% CI [-0.893, -0.511],  $r = 0.39$ . In contrast, the novelty manipulation did not impact surprise in Experiment 1,  $b = -0.08$ ,  $se = 0.08$ ,  $t(223) = -1.04$ ,  $p = 0.30$ , 95% CI [-0.246, 0.076],  $r = 0.07$ , or Experiment 2,  $b = 0.02$ ,  $se = 0.10$ ,  $t(298) = 0.25$ ,  $p = 0.80$ , 95% CI [-0.166, 0.215],  $r = 0.01$ . Of note, participants' ratings of novelty and their ratings of surprise were not positively correlated in either study as would be expected from semantic-free information theory. To the contrary, a small negative correlation was observed in both Experiment 1,  $r(224) = -0.24$ ,  $p < 0.001$ , and Experiment 2,  $r(299) = -0.18$ ,  $p = 0.002$ . An alternative analysis strategy to examine the impact of belief consistency would be to inspect interactions between information and belief directions. See Supplementary Data 1 for such analyses on each of the outcomes.

### *Belief-consistent information more valued and shared*

In Experiment 2, belief-consistent information was rated as more valuable than belief-inconsistent information,  $b = 0.20$ ,  $se = 0.07$ ,  $t(298) = 2.77$ ,  $p = 0.006$ , 95% CI [0.058, 0.341],  $r = 0.16$ . In contrast, the novelty manipulation had no significant effect on perceived value,  $b = 0.03$ ,  $se = 0.07$ ,  $t(298) = 0.43$ ,  $p = 0.67$ , 95% CI [-0.110, 0.172],  $r = 0.02$ . This preference for belief-consistent information parallels the pattern of actual sharing decisions (see Fig. 4).

Because our information sharing outcome variable is a count of the amount of update information that was shared, these analyses were conducted using ordinal logistic regression. Recall that participants were exposed to two pieces of information within the update, such that they could decide to transmit zero, one, or two pieces of update information to the editor. Participants were more likely to share information that was belief consistent rather than belief inconsistent in both Experiment 1,  $b = 0.40$ ,  $se = 0.13$ ,  $t(222) = 3.07$ , 95% CI [0.147, 0.662], *Odds Ratio* (OR) = 1.50, and Experiment 2,  $b = 0.37$ ,  $se = 0.12$ ,  $t(297) = 3.05$ , 95% CI [0.132, 0.602], OR = 1.44. The effect of the novelty manipulation on the amount of update information shared did not reach significance



**Figure 4.** Effect of update information consistency on the amount of update information shared. Each bar indicates the mean amount of update information shared by participants. Error bars represent standard error.

in Experiment 1,  $b = -0.01$ ,  $se = 0.13$ ,  $t(222) = -0.10$ , 95% CI  $[-0.269, 0.243]$ ,  $OR = 0.99$ , or in Experiment 2,  $b = -0.09$ ,  $se = 0.12$ ,  $t(297) = -0.76$ , 95% CI  $[-0.327, 0.144]$ ,  $OR = 0.91$ .

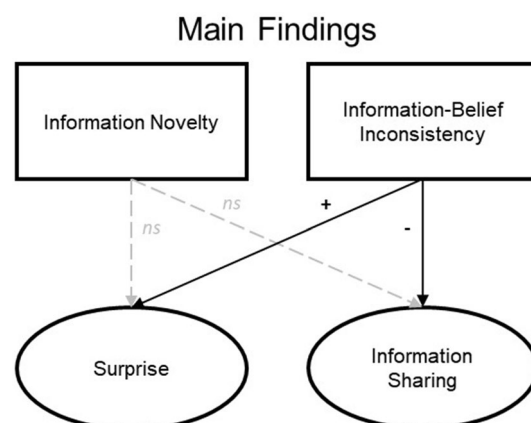
#### Novelty manipulations undetected

To explore why our novelty manipulation failed to impact sharing intentions, we tested whether *subjective* novelty ratings for the update information differed across conditions. The novelty manipulation had no significant effect on perceived novelty in Experiment 1,  $b = 0.08$ ,  $se = 0.10$ ,  $t(223) = 0.73$ ,  $p = 0.47$ , 95% CI  $[-0.129, 0.280]$ ,  $r = 0.05$ , or in Experiment 2,  $b = -0.03$ ,  $se = 0.08$ ,  $t(298) = -0.43$ ,  $p = 0.67$ , 95% CI  $[-0.190, 0.122]$ ,  $r = 0.02$ , suggesting that participants did not consistently identify information they had not seen before as being more novel than information they had seen before. Moreover, the belief consistency manipulation had no significant impact on perceived novelty in Experiment 1,  $b = 0.13$ ,  $se = 0.10$ ,  $t(223) = 1.28$ ,  $p = 0.20$ , 95% CI  $[-0.072, 0.337]$ ,  $r = 0.09$ , or in Experiment 2,  $b = 0.15$ ,  $se = 0.08$ ,  $t(298) = 1.91$ ,  $p = 0.06$ , 95% CI  $[-0.005, 0.309]$ ,  $r = 0.11$ . Thus, it seems that influences of belief consistency are unique to surprise rather than uniformly influencing both surprise and perceived novelty.

## Discussion

In the field and in the lab, we found that information's consistency with prior beliefs reliably predicts information sharing, in line with the context-sensitive perspective (and contrary to the semantic-free perspective). Our experimental findings also suggest that novelty and surprise are distinct concepts with unique consequences for information sharing (see Fig. 5). Unexpectedly, although our Twitter analyses were consistent with information novelty predicting sharing, our experiments that provide more control and a more precise manipulation of novelty offered no evidence that novelty was related to sharing.

Our experimental findings diverged in notable ways from what semantic-free information theory would predict<sup>1,2</sup>. First, novel (previously unencountered) information was not shared more than information that was



**Figure 5.** Conceptual depiction of the main findings from Experiments 1 and 2.

not novel (previously encountered). Furthermore, *less* surprising information was most likely to be transmitted, directionally opposite to assumptions in information theory. The only positive predictor of sharing was information-belief consistency, and such a preference for belief-consistent information might be even stronger when information pertains to more closely-held beliefs (e.g., political views<sup>37,48</sup>). These results align with other work demonstrating the relevance of contextual factors to information spread within networks<sup>51</sup>.

Also of interest was participants' inability to distinguish between information that was objectively novel versus not novel in their ratings of information novelty or surprise. Our novelty manipulation was unequivocal; information was either previously encountered or not. This insensitivity to objective novelty evident in participants' subjective appraisals might explain why the novelty manipulation failed to impact sharing or perceived value. On the other hand, participants' levels of surprise in response to belief-consistent versus belief-inconsistent information showed that participants did react differently to information based on its belief consistency. Information inconsistent with prior beliefs was perhaps more surprising because it was deemed less likely to be true<sup>52</sup>. As such, it seems that people might react more to information's belief consistency and base their sharing decisions more on that factor than on the information's novelty. Notably, our belief consistency manipulation had no impact on perceived novelty, following prior research distinguishing these constructs<sup>25–27</sup>.

The large-scale observational data have a major limitation: Twitter's recommendation algorithm is a black box. We know that Twitter promotes content based on recency and engagement<sup>53</sup>, and it is at least possible that it also prioritizes novelty. The details of the algorithm, however, are not publicly available. This means that we have no way of assessing the extent to which the association between novelty and sharing in our Twitter data is due to individual preferences versus Twitter's algorithms. This is an important limitation, but it does not undermine our claim that sharing predictions based on semantic-free information theory are incomplete. First, we note that the extant literature provides ample evidence that novelty promotes sharing<sup>11–16,47</sup>. Given this, it is likely that at least some of the effects of novelty observed here are, in fact, attributable to individual preferences. Second, and more importantly, we demonstrate that belief-inconsistency (which is correlated with surprise) is a significant negative predictor of sharing regardless of the level of information novelty—regardless of whether it is a by-product of individual preference or algorithmic recommendation.

The experiments have limitations, too. First, they both utilized unfamiliar topics about which participants should have had relatively weak (if any) pre-existing beliefs. This was desirable in that it allowed us to better control the variables of interest (i.e., the novelty of new information and its consistency with beliefs established in the prior phase), ensuring that factors outside of our experimental procedures would not interfere with the results. However, the extent to which the present experimental findings generalize to contexts where people do hold strong prior beliefs, such as the political (as was examined in our Twitter analyses) or health<sup>54</sup> domains, remains unclear. Future research should examine these experimental results using topics where pre-existing beliefs exist to assess the present findings' generalizability to those contexts. Similarly, because the information sharing tasks in the present experiments were hypothetical, there were no real consequences to the participants or others associated with what information was or was not shared. Real world information sharing is often consequential, especially if the shared information is incorrect or leads to incorrect conclusions. Though our tasks were designed to prompt participants to consider such consequences, they were nevertheless hypothetical. As such, future research should continue to examine information sharing in settings where decisions can have real consequences, similar to our analyses of belief consistency in Twitter retweets. The experimental scenario also likely minimized participants' self-presentational and affiliative motives, both core drivers of information sharing<sup>46,47,55,56</sup>. Research should attempt to replicate our findings with these concerns made salient.

Taken as a whole, this work demonstrates that when modeling information sharing, we must distinguish between novelty and surprise. Using Twitter data, we show that the people sharing tweets are more likely to agree than disagree with the political leanings of the tweets regardless of the information's novelty. The pair of experiments further distinguish between novelty in surprise by showing that a novelty manipulation was not related to surprise and that (lack of) surprise was associated with sharing (of belief-consistent information), whereas novelty was not associated with information sharing. Surprise, in the form of belief consistency, is the best predictor of information sharing, but in a direction opposite to that assumed by semantic-free information theory (i.e., less surprise was associated with more sharing).

More generally, these results suggest that context-sensitive predictions are more accurate than those based on semantic-free information theoretic principles. Indeed, recent research documents that false partisan news propagates more quickly in ideologically segregated networks, seemingly because the implausibility of ideologically-inconsistent false news discourages diffusion in mixed networks<sup>57</sup>. This accords with other work on the propagation of misinformation in polarized “echo chambers” online<sup>58–61</sup>. One important ingredient for constructing such models thus appears to be the judicious combination of information theoretic concepts with a principled approach for studies of subjective perceptions that allow for capturing more relevant variables in the context of interest.

The present research opens a number of doors for future investigation. First, research should examine whether other factors separate from, but correlated with, information novelty predict information sharing. Namely, our Twitter analysis is consistent with the idea that information that is ‘fresh’ (e.g., released close in time to a focal events' occurrence) is more likely to be shared than less fresh information. If novelty in the form of “freshness” reliably predicts sharing, it would be interesting to consider what aspects of information freshness might lead to greater sharing. One possibility is that more recent events, whether or not they are particularly novel to perceivers, are more likely to be current topics of conversation than older events. If so, people might be motivated to engage with fresher tweets to stay included in current conversations<sup>62</sup>. People also might be more likely to see fresher tweets that attract more engagement than less fresh tweets, amplifying the reach of those fresher tweets. Testing such possibilities in an experimental context would be beneficial.

These theoretical insights also have practical implications. For example, the findings involving belief consistency could suggest that framing information as consistent with the prior beliefs of target populations might increase its likelihood of retransmission. For instance, broadcasted messages (e.g., PSAs) might benefit from framing arguments as consistent with personally important moral foundations<sup>63</sup>, values<sup>64</sup>, or political ideologies<sup>65</sup> of target audiences. Likewise, the current findings suggest that it might not be safe to assume that novel information will be spread by recipients. This might be especially important in contexts where information is rapidly changing, such as in crisis situations<sup>66</sup>, where the novelty of new guidance might not be sufficient to ensure its spread. In such cases, additional steps, such as framing the guidance to be as consistent with prior-held beliefs as possible, might be beneficial. Further research should explore these possibilities. Our findings are also consistent with how pre-existing beliefs hinder corrections of political misinformation<sup>67,68</sup>. Whether corrections are also less effective at attenuating the *transmission* of belief-consistent misinformation is a ripe area for contemporary research<sup>36,40,69–72</sup>.

Complementing observational and experimental work, agent-based models<sup>73</sup> offer a lens through which to evaluate factors that might bolster interventions to curtail the spread of (mis)information at scale<sup>74–78</sup>. That is, such models allow for clear manipulations of individual-level factors (e.g., prior beliefs and information exposure) to explore macro-level outcomes (e.g., large-scale information diffusion). Social media field experiments are another complementary direction for future work that can lay the foundation for real-world interventions<sup>79,80</sup>. These approaches could help discern whether corrections or inoculations targeting people for whom circulating misinformation is likely to be belief consistent are ideal. For example, such interventions might only be worthwhile to the extent that messages emphasize consistency with prior beliefs while debunking the specific piece of misinformation.

Overall, the present findings suggest that the meaning of information can modulate the typical effects of novelty and surprise expected from a straightforward application of semantic-free information theory. Specifically, we found that information-belief consistency (a key determinant of surprise) and novelty have distinct influences on sharing intentions. Belief consistency is associated with low levels of surprise but high levels of sharing. As such, our conclusion is that care must be exercised in applying information theoretic principles when sharing occurs in contexts where people hold pre-existing beliefs relevant to information they receive. From a big picture perspective, it seems important to consider the semantics of information when approaching real-world information sharing problems.

## Methods

### Twitter analyses methods

#### Data

Using the Twitter API, we downloaded 223 tweets and the associated metadata, including when each message was posted and the number of times it was retweeted ( $M = 339.68$ ,  $SD = 1022.00$ ). Each tweet pertained to a distinct, discrete event that occurred at a specific point in time. The tweets were posted by relatively non-partisan news organizations such as Reuters, the Associated Press, Newsweek, and Axios. These organizations were selected using ratings from AllSides, an independent organization that specializes in assessing media bias across the political landscape<sup>81</sup>. All news outlets included in the present study were deemed politically neutral at the time of data collection. This was done to make it unlikely that the followers of the respective accounts would be systematically biased in the liberal or conservative directions. Third, we sought events about which non-partisan news sources tweeted multiple times.

All 223 tweets were used for the novelty analyses. For the belief consistency analyses, we also downloaded user IDs of the thousands of users who retweeted the selected messages and all those whom they followed. That is, for each of roughly 100 tweets (43 in the first analysis and 64 in the second), each retweeter (~300 per tweet) followed on average approximately 2000 accounts (that were used to determine the ideology of each retweeter). Thus, the final datasets for the belief consistency analyses included over sixty million data points.

#### Sharing

For analyses addressing novelty and sharing, we operationalized information sharing as the number of retweets a given tweet received. When examining relations between belief consistency and sharing, however, we classified the ideological leanings of people who retweeted clearly pro-liberal versus pro-conservative tweets.

#### Novelty

We used time, in minutes, between when the tweet being evaluated was posted and when the first tweet about the same event was posted as a proxy measure for information novelty. This time variable was standardized for use in the analysis.

#### Ideological consistency

To assess the ideological consistency of the tweets, we compared the ideological slant of the post to the ideological orientation of the user sharing it. If the two ideologies matched, we labeled the tweet ideologically consistent. To assess the ideological slant of the tweets, two members of the research team coded by consensus whether each tweet was liberal- or conservative-aligned. That is, we examined whether the message espoused or promoted liberal or conservative positions (see Supplementary Data 2 for examples; the full list of Tweet IDs is available on OSF<sup>82</sup>). Ideological consistency was based on the political slant of the tweet, not the event.

To assess retweeters' ideological orientation, we adapted a method from prior research that builds on the assumption that people tend to follow politicians who share their political ideology<sup>42</sup>. For example, those who follow the most liberal/conservative members of the U.S. Congress tend to hold more ideologically extreme views.



First, we randomly selected eight events (4 liberal-aligned and 4 conservative-aligned) from the sample of 223 tweets used in the novelty analysis and extracted all the tweets pertaining to that event (43 tweets total). We also performed an identical analysis on a second, separate sample of 64 tweets that met the same criteria as the first.

Next, we obtained estimates of each U.S. Senate and House member's ideology from govtrack.us, a non-partisan organization that tracks congressional events and member voting activity<sup>83</sup>. These estimates are derived from a member's voting behavior and a comparison between that behavior and the behavior of other members of congress. Ideology estimates range from 0 (most liberal) to 1 (most conservative). We also collected both the professional and personal Twitter user IDs of all US Senate and House members.

Finally, we determined whether each retweeter followed one or more U.S. House and/or Senate members. If so, we assigned that retweeter with an ideology score that was the simple average of the ideology scores of the politicians whom they were following. Those users who did not follow a member of Congress were excluded from the analyses. This resulted in 4413 and 5724 retweeters being included in the first and second analyses, respectively.

## Experiment methods

### *Participants and design*

Two hundred ninety-four Introduction to Psychology students enrolled in Ohio State University's Research Experience Program participated online in Experiment 1 in exchange for partial course credit. Forty-eight participants were excluded from analysis due to being duplicate responses or failing to complete all of the key measures, resulting in an analyzed sample of 226 (Age:  $M = 18.90$ ,  $SD = 1.60$ ; Gender: 38.9% male, 59.3% female, 1.8% other or no response). Three hundred four workers on Amazon's Mechanical Turk were recruited to participate in Experiment 2 in exchange for \$1.50. Of these, 3 participants were excluded from analyses using the same criteria as Experiment 1, resulting in an analyzed sample of 301 (Age  $M = 39.45$ ,  $SD = 11.71$ ; Gender: 52.3% male, 46.7% female, 1% non-binary).

Both studies employed a 2(Update Information Novelty: Not Novel vs. Novel) X 2(Update Information Consistency with Initial Beliefs: Consistent vs. Inconsistent) X 2(Belief Direction: Riskiness is Good for Firefighting/Country Should be Allowed to Join the EU vs. Riskiness is Bad for Firefighting/Country Should Not be Allowed to Join the EU) between-subjects design.

### *Procedure*

After providing informed consent, participants in Experiment 1 were told they would take on the role of an editor's assistant at a news organization. Their goal was to decide to transmit as much information to the editor as necessary to understand the conclusion drawn from the total evidence uncovered by reporters. They read an initial interview transcript under the premise that they would soon receive an update from the reporter. The interview was reportedly conducted with a fire chief and concerned the merits of risky behavior for effective firefighting. With this transcript, participants' beliefs about the relation between riskiness and firefighting success were manipulated. Participants were randomly assigned to see a transcript in which the fire chief, as well as most embedded supporting information (5 of the total 7 pieces), suggested that riskiness was either beneficial or detrimental. The other 2 pieces of information were in opposition to the primary position taken by the chief. Our experimental paradigm drew from prior studies involving newly established beliefs<sup>49,50</sup>.

Participants then reported their belief about this relation on a continuous scale before viewing the reporter's "update." Within the update, another source presented two pieces of evidence either supporting or opposing the conclusion from the prior interview. As such, the consistency of the update information with participants' initial beliefs was manipulated between subjects. Additionally, this supporting or opposing information was either the same as information participants had seen in the transcript (familiar information) or different from the information in the transcript (novel information).

After again reporting their belief, participants rated each piece of information included within the update on the degree to which it was novel and surprising. They were then asked to report which information they deemed necessary to transmit to the newspaper editor.

Study 2 employed a largely identical procedure, save a few key distinctions. Chiefly, the reporter's topic was modified to concern whether or not an unnamed country should be allowed to join the European Union (adapted from prior research<sup>84</sup>). Participants also completed a four-item measure of "value" for each piece of update information before making their transmission decisions. Lastly, the ratings of novelty and surprise occurred following, rather than prior to, the sharing phase to ensure that our findings were not due to increased salience of these features at the time of sharing.

### *Dependent variables*

**Novelty Ratings.** In both studies, participants' assessments of how novel they found each piece of information in the update were assessed using three items. For pieces of information explaining the link between riskiness and firefighting success, they were asked to indicate the extent to which they had encountered that explanation before (1—I've Never Encountered This Explanation Before to 7—I've Definitely Encountered this Explanation Before), how familiar they were with that explanation (1—Not at All Familiar to 7—Extremely Familiar), and how often they had seen the explanation before (1—I Have Never Heard or Seen This Explanation to 7—I Have Heard or Seen This Explanation Very Often). Responses to each of these three items were averaged together for each piece of information, and then these two indexes were averaged together to create a single index of perceived novelty of the update information. The composite scales demonstrated adequate reliability in both Experiment 1 ( $\alpha = 0.90$ ,  $M = 4.29$ ,  $SD = 1.56$ ) and Experiment 2 ( $\alpha = 0.90$ ,  $M = 5.07$ ,  $SD = 1.38$ ).

**Surprise Ratings.** In both studies, participants indicated how surprising they found the update information to be on three items. Participants were asked to what extent they found each of the two explanations to be

surprising (1—Not at All Surprising to 7—Very Surprising) and unexpected (1—Not at All Unexpected to 7—Very Unexpected), and whether each explanation caught them off guard (1—Not at All to 7—Very Much). Responses to the three items were averaged together for each of the pieces of information and the two indexes were averaged together to create a single index of surprise. The composite scales demonstrated adequate reliability in both Experiment 1 ( $\alpha = 0.90$ ,  $M = 2.94$ ,  $SD = 1.37$ ) and Experiment 2 ( $\alpha = 0.96$ ,  $M = 3.37$ ,  $SD = 1.82$ ).

**Belief ratings and belief rating shifts.** In Experiment 1, participants' beliefs about whether riskiness is good or bad for firefighting were assessed using one item immediately following the transcript. This item asked participants to provide their estimate of the direction and strength of the relation between success as a firefighter and risk taking. Participants were reminded that a positive relation means that high riskiness is associated with success as a firefighter, and a negative relation means that high riskiness is associated with failure as a firefighter. Responses were given on a single slider scale (–50—Highly Negative Relationship to 50—Highly Positive Relationship). Beliefs were assessed again using the same scale after participants read the update information. To assess shifts in beliefs from before exposure to the update information to after, a difference score was calculated by subtracting the post-update information belief ratings from the initial belief ratings. As such, positive values of the difference score indicated that beliefs shifted in the direction of believing that riskiness is good for firefighting, whereas negative values corresponded to shifts in beliefs in the direction of believing that riskiness is bad for firefighting ( $M = -0.06$ ,  $SD = 35.33$ ).

In Experiment 2, similar items were used except they assessed beliefs about whether the country should or should not be allowed to join the EU (–50—The Country Should NOT be Allowed to Join the EU to 50—The Country SHOULD be Allowed to Join the EU). A difference score was computed in the same fashion. Positive values indicated shifts in beliefs in the direction of believing that the country should be allowed to join, whereas negative values indicated shifts in the direction of believing that the country should not be allowed to join ( $M = -3.12$ ,  $SD = 23.74$ ).

**Perceived value of the update information.** In Experiment 2, four items assessed how valuable participants found each of the two pieces of information within the update to be. Participants reported the extent to which they thought the information was worth sharing (1—Not at All Worth Sharing to 7—Extremely Worth Sharing), valuable (1—Not at All Valuable to 7—Very Valuable), and important to their overall conclusion, (1—Not at All Important to 7—Very Important). Participants also rated whether they would include the information in their report to their editor (1—Definitely No to 7—Definitely Yes). All responses to the four items for each piece of information were averaged together and then the resulting two indexes were averaged to create a single overall index of perceived value. The composite scale demonstrated adequate reliability ( $\alpha = 0.94$ ,  $M = 5.62$ ,  $SD = 1.26$ ).

**Intention to share information.** In both studies, participants were presented with a list of all the information they had encountered from both the initial transcript and the update. Participants were asked to indicate which pieces of information they would like to include with their report to their editor by dragging each piece of information from the list into a box labeled “include” or a box labeled “exclude.” We assessed participants' desire to share the update information by counting how many of the pieces of information from the update they placed in the include box. Participants could include both explanations from the update, just one explanation, or neither of them in Experiment 1 ( $M = 1.41$ ,  $SD = 0.71$ ) and Experiment 2 ( $M = 1.49$ ,  $SD = 0.76$ ).

## Ethics

Human subjects research was determined exempt from full review and approved by The Ohio State University Human Research Protection Program (HRPP) under protocol number 2021E1208 (Federalwide Assurance (FWA) for the Protection of Human Subjects Number: 00006378). All research was performed in accordance with relevant guidelines and regulations. Informed consent was obtained from all human research participants.

## Data availability

Twitter and experimental data files have been made available in the Open Science Framework [[https://osf.io/wy3hp/?view\\_only=b4f98d8ddfd746a1a71df412ee829a6a](https://osf.io/wy3hp/?view_only=b4f98d8ddfd746a1a71df412ee829a6a)]<sup>83</sup>. In accordance with content redistribution policies for Twitter Content, we are restricted to sharing only Tweet IDs and User IDs.

## Code availability

R code used in analyses has been made available in the Open Science Framework [[https://osf.io/wy3hp/?view\\_only=b4f98d8ddfd746a1a71df412ee829a6a](https://osf.io/wy3hp/?view_only=b4f98d8ddfd746a1a71df412ee829a6a)]<sup>83</sup>.

Received: 9 November 2023; Accepted: 1 March 2024

Published online: 13 March 2024

## References

- Cover, T. M. & Thomas, J. A. *Elements of Information Theory*. <https://doi.org/10.1002/0471200611> (Wiley, 1991).
- Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> (1948).
- Barto, A., Mirolli, M. & Baldassarre, G. Novelty or surprise?. *Front. Psychol.* **4**, 907. <https://doi.org/10.3389/fpsyg.2013.00907> (2013).
- Palm, G. *Novelty, Information and Surprise*. <https://doi.org/10.1007/978-3-642-29075-6> (Springer, 2023).
- González-Bailón, S. & Wang, N. Networked discontent: The anatomy of protest campaigns in social media. *Soc. Netw.* **44**, 95–104. <https://doi.org/10.1016/j.socnet.2015.07.003> (2016).
- Tremayne, M. Anatomy of protest in the digital era: A network analysis of Twitter and occupy wall street. *Soc. Mov. Stud.* **13**, 110–126. <https://doi.org/10.1080/14742837.2013.830969> (2014).
- Purohit, H., Ruan, Y., Fuhr, D., Parthasarathy, S. & Sheth, A. On understanding the divergence of online social group discussion. *ICWSM 8*, 396–405. <https://doi.org/10.1609/icwsml.14557> (2014).

8. Shi, J., Poorisat, T. & Salmon, C. T. The use of social networking sites (SNSs) in health communication campaigns: Review and recommendations. *Health Commun.* **33**, 49–56. <https://doi.org/10.1080/10410236.2016.1242035> (2018).
9. Van der Linden, S. Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nat. Med.* **28**, 460–467. <https://doi.org/10.1038/s41591-022-01713-6> (2022).
10. Compton, J., van der Linden, S., Cook, J. & Basol, M. Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Soc. Pers. Psychol. Compass* **15**, e12602. <https://doi.org/10.1111/spc3.12602> (2021).
11. Kim, H. S., Lee, S., Cappella, J. N., Vera, L. & Emery, S. Content characteristics driving the diffusion of antismoking messages: Implications for cancer prevention in the emerging public communication environment. *JNCI Monographs* **182–187**, 2013. <https://doi.org/10.1093/jncimonographs/ltg018> (2013).
12. Photiou, A., Nicolaides, C. & Dhillon, P. S. Social status and novelty drove the spread of online information during the early stages of COVID-19. *Sci. Rep.* **11**, 1–7. <https://doi.org/10.1038/s41598-021-99060-y> (2021).
13. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151. <https://doi.org/10.1126/science.aap9559> (2018).
14. Chen, X., Pennycook, G. & Rand, D. What makes news sharable on social media?. *JQD:DM* **3**, 1–27. <https://doi.org/10.51685/jqd.2023.007> (2023).
15. Berger, J. & Milkman, K. L. What makes online content viral?. *J. Mark. Res.* **49**, 192–205. <https://doi.org/10.1509/jmr.10.0353> (2012).
16. Altay, S., De Araujo, E. & Mercier, H. “If this account is true, it is most enormously wonderful”: Interestingness-if-true and the sharing of true and false news. *Digit. Journal.* **10**, 373–394. <https://doi.org/10.1080/21670811.2021.1941163> (2022).
17. Kumari, R., Ashok, N., Ghosal, T. & Ekbal, A. What the fake? Probing misinformation detection standing on the shoulder of novelty and emotion. *Inf. Process. Manag.* **59**, 102740. <https://doi.org/10.1016/j.ipm.2021.102740> (2022).
18. Loewenstein, J. & Heath, C. The repetition-break plot structure: A cognitive influence on selection in the marketplace of ideas. *Cogn. Sci.* **33**, 1–19. <https://doi.org/10.1111/j.1551-6709.2008.01001.x> (2009).
19. Baker, S. M. & Petty, R. E. Majority and minority influence: Source-position imbalance as a determinant of message scrutiny. *J. Pers. Soc. Psychol.* **67**, 5–19. <https://doi.org/10.1037/0022-3514.67.1.5> (1994).
20. Ziegler, R., Diehl, M. & Ruther, A. Multiple source characteristics and persuasion: Source inconsistency as a determinant of message scrutiny. *Pers. Soc. Psychol. Bull.* **28**, 496–508. <https://doi.org/10.1177/0146167202287007> (2002).
21. Tribus, M. *Thermodynamics and Thermostatistics: An Introduction to Energy, Information and States of Matter, with Engineering Applications* (Van Nostrand, 1961).
22. Itti, L. & Baldi, P. A principled approach to detecting surprising events in video. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (eds Schmid, C. et al.) 631–637 (IEEE, Piscataway, 2005). <https://doi.org/10.1109/CVPR.2005.40>
23. Itti, L. & Baldi, P. Bayesian surprise attracts human attention. *Vis. Res.* **49**, 1295–1306. <https://doi.org/10.1016/j.visres.2008.09.007> (2009).
24. Bullock, J. G. Partisan bias and the Bayesian ideal in the study of public opinion. *J. Politics* **71**, 1109–1124. <https://doi.org/10.1017/S0022381609090914> (2009).
25. Bardin, B., Vidal, P., Facca, L., Dumas, R. & Perrissol, S. The effect of information quality evaluation on selective exposure in informational cognitive dissonance: The role of information novelty. *Int. Rev. Soc. Psychol.* **31**, 21. <https://doi.org/10.5334/irsp.173> (2018).
26. Brannon, S. M., Sacchi, D. L. & Gawronski, B. (In)consistency in the eye of the beholder: The roles of warmth, competence, and valence in lay perceptions of inconsistency. *J. Exp. Soc. Psychol.* **70**, 80–94. <https://doi.org/10.1016/j.jesp.2016.12.011> (2017).
27. Vachon, F., Hughes, R. W. & Jones, D. M. Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *JEP:LMC* **38**, 164–177. <https://doi.org/10.1037/a0025054> (2012).
28. Yang, Q., Tufts, C., Ungar, L., Guntuku, S. & Merchant, R. To retweet or not to retweet: Understanding what features of cardiovascular tweets influence their retransmission. *J. Health Commun.* **23**, 1026–1035. <https://doi.org/10.1080/10810730.2018.1540671> (2018).
29. Kim, H. S. Attracting views and going viral: How message features and news-sharing channels affect health news diffusion. *J. Commun.* **65**, 512–534. <https://doi.org/10.1111/jcom.12160> (2015).
30. Taber, C. S. & Lodge, M. Motivated skepticism in political information processing. *Am. J. Political Sci.* **50**, 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x> (2006).
31. Kim, J. H., Bock, G. W., Sabherwal, R. & Kim, H. M. Why do people spread online rumors? An empirical study. *APJIS* **29**, 591–614. <https://doi.org/10.14329/apjis.2019.29.4.591> (2019).
32. Shin, J. & Thorson, K. Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *J. Commun.* **67**, 233–255. <https://doi.org/10.1111/jcom.12284> (2017).
33. Bowyer, B. & Kahne, J. Motivated circulation: How misinformation and ideological alignment influence the circulation of political content. *Int. J. Commun.* **13**, 5791–5815 (2019).
34. Buchanan, T. Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS ONE* **15**, e0239666. <https://doi.org/10.1371/journal.pone.0239666> (2020).
35. Giese, H., Neth, H., Moussaïd, M., Betsch, C. & Gaissmaier, W. The echo in flu-vaccination echo chambers: Selective attention trumps social influence. *Vaccine* **38**, 2070–2076. <https://doi.org/10.1016/j.vaccine.2019.11.038> (2020).
36. Pennycook, G., Bear, A., Collins, E. T. & Rand, D. G. The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Manage. Sci.* **66**, 4944–4957. <https://doi.org/10.1287/mnsc.2019.3478> (2020).
37. Pennycook, G. et al. Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595. <https://doi.org/10.1038/s41586-021-03344-2> (2021).
38. Dechêne, A., Stahl, C., Hansen, J. & Wänke, M. The truth about the truth: A meta-analytic review of the truth effect. *Pers. Soc. Psychol. Rev.* **14**, 238–257. <https://doi.org/10.1177/1088868309352251> (2010).
39. Pennycook, G., Cannon, T. D. & Rand, D. G. Prior exposure increases perceived accuracy of fake news. *J. Exp. Psychol. Gen.* **147**, 1865–1880. <https://doi.org/10.1037/xge0000465> (2018).
40. Efron, D. A. & Raj, M. Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychol. Sci.* **31**, 75–87. <https://doi.org/10.1177/0956797619887896> (2020).
41. Vellani, V., Zheng, S., Erceci, D. & Sharot, T. The illusory truth effect leads to the spread of misinformation. *Cognition* **236**, 105421. <https://doi.org/10.1016/j.cognition.2023.105421> (2023).
42. Golbeck, J. & Hansen, D. A method for computing political preference among Twitter followers. *Soc. Netw.* **36**, 177–184. <https://doi.org/10.1016/j.socnet.2013.07.004> (2014).
43. Jerit, J. & Barabas, J. Partisan perceptual bias and the information environment. *J. Politics* **74**, 672–684. <https://doi.org/10.1017/S0022381612000187> (2012).
44. Funk, C. & Hefferon, M. U.S. public views on climate and energy. <https://www.pewresearch.org/science/2019/11/25/u-s-public-views-on-climate-and-energy/> (2019).

45. Baumeister, R. F. A self-presentational view of social phenomena. *Psychol. Bull.* **91**, 3–26. <https://doi.org/10.1037/0033-2909.91.1.3> (1982).
46. Berger, J. Word of mouth and interpersonal communication: A review and directions for future research. *J. Consum. Psychol.* **24**, 586–607. <https://doi.org/10.1016/j.jcps.2014.05.002> (2014).
47. Cappella, J. N., Kim, H. S. & Albarracín, D. Selection and transmission processes for information in the emerging media environment: Psychological motives and message characteristics. *Media Psychol.* **18**, 396–424. <https://doi.org/10.1080/15213269.2014.941112> (2015).
48. Pretus, C. *et al.* The role of political devotion in sharing partisan misinformation and resistance to fact-checking. *J. Exp. Psychol. Gen.* **152**, 3116–3134. <https://doi.org/10.1037/xge0001436> (2023).
49. Anderson, C. A., Lepper, M. R. & Ross, L. Perseverance of social theories: The role of explanation in the persistence of discredited information. *J. Pers. Soc. Psychol.* **39**, 1037–1049. <https://doi.org/10.1037/h0077720> (1980).
50. Anderson, C. A. Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *J. Exp. Soc. Psychol.* **19**, 93–108. [https://doi.org/10.1016/0022-1031\(83\)90031-8](https://doi.org/10.1016/0022-1031(83)90031-8) (1983).
51. Vedula, N., Parthasarathy, S. & Shalin, V. L. Predicting trust relations among users in a social network: The role of influence, cohesion and valence. In *Proceedings of ACM SIGKDD WISDOM (2016)* (2016).
52. Metzger, M. J., Hartsell, E. H. & Flanagan, A. J. Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Commun. Res.* **47**, 3–28. <https://doi.org/10.1177/0093650215613136> (2020).
53. Twitter's recommendation algorithm. [https://blog.twitter.com/engineering/en\\_us/topics/open-source/2023/twitter-recommendation-algorithm](https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm) (2023)
54. Wang, R., He, Y., Xu, J. & Zhang, H. Fake news or bad news? Toward an emotion-driven cognitive dissonance model of misinformation diffusion. *Asian J. Commun.* **30**, 317–342. <https://doi.org/10.1080/01292986.2020.1811737> (2020).
55. Lee, C. S. & Ma, L. News sharing in social media: The effect of gratifications and prior experience. *Comput. Hum. Behav.* **28**, 331–339. <https://doi.org/10.1016/j.chb.2011.10.002> (2012).
56. Duffy, A., Tandoc, E. & Ling, R. Too good to be true, too good not to share: The social utility of fake news. *Inf. Commun. Soc.* **23**, 1965–1979. <https://doi.org/10.1080/1369118X.2019.1623904> (2020).
57. Stein, J., Keuschnigg, M. & van de Rijt, A. Network segregation and the propagation of misinformation. *Sci. Rep.* **13**, 917. <https://doi.org/10.1038/s41598-022-26913-5> (2023).
58. Del Vicario, M. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 554–559. <https://doi.org/10.1073/pnas.1517441113> (2016).
59. Bessi, A. *et al.* Homophily and polarization in the age of misinformation. *Eur. Phys. J. Spec. Top.* **225**, 2047–2059. <https://doi.org/10.1140/epjst/e2015-50319-0> (2016).
60. Törnberg, P. Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE* **13**, e0203958. <https://doi.org/10.1371/journal.pone.0203958> (2018).
61. Iandoli, L., Primario, S. & Zollo, G. The impact of group polarization on the quality of online debate in social media: A systematic literature review. *Technol. Forecast. Soc. Change* **170**, 120924. <https://doi.org/10.1016/j.techfore.2021.120924> (2021).
62. Baumeister, R. F. & Leary, M. R. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* **117**, 497–529. <https://doi.org/10.1037/0033-2909.117.3.497> (1995).
63. Day, M. V., Fiske, S. T., Downing, E. L. & Trail, T. E. Shifting liberal and conservative attitudes using moral foundations theory. *Pers. Soc. Psychol. Bull.* **40**, 1559–1573. <https://doi.org/10.1177/0146167214551152> (2014).
64. Shen, F. & Edwards, H. H. Economic individualism, humanitarianism, and welfare reform: A value-based account of framing effects. *J. Commun.* **55**, 795–809. <https://doi.org/10.1111/j.1460-2466.2005.tb03023.x> (2005).
65. Luong, K. T., Garrett, R. K. & Slater, M. D. Promoting persuasion with ideologically tailored science messages: A novel approach to research on emphasis framing. *Sci. Commun.* **41**, 488–515. <https://doi.org/10.1177/1075547019862559> (2019).
66. Janssen, M., Lee, J., Bharosa, N. & Cresswell, A. Advances in multi-agency disaster management: Key elements in disaster research. *Inf. Syst. Front.* **12**, 1–7. <https://doi.org/10.1007/s10796-009-9176-x> (2010).
67. Walter, N., Cohen, J., Holbert, R. L. & Morag, Y. Fact-checking: A meta-analysis of what works and for whom. *Polit. Commun.* **37**, 350–375. <https://doi.org/10.1080/10584609.2019.1668894> (2020).
68. Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N. & Cook, J. Misinformation and its correction: Continued influence and successful debiasing. *Psychol. Sci. Public Interest* **13**, 106–131. <https://doi.org/10.1177/1529100612451018> (2012).
69. Friggeri, A., Adamic, L., Eckles, D. & Cheng, J. Rumor cascades. *ICWSM* **8**, 101–110. <https://doi.org/10.1609/icwsm.v8i1.14559> (2014).
70. Mosleh, M., Martel, C., Eckles, D. & Rand, D. Perverse downstream consequences of debunking: Being corrected by another user for posting false political news increases subsequent sharing of low quality, partisan, and toxic content in a Twitter field experiment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* 1–13 (ACM, New York, 2021). <https://doi.org/10.1145/3411764.3445642>
71. Chen, S., Xiao, L. & Kumar, A. Spread of misinformation on social media: What contributes to it and how to combat it. *Comput. Hum. Behav.* **141**, 107643. <https://doi.org/10.1016/j.chb.2022.107643> (2022).
72. Freiling, I., Krause, N. M., Scheufele, D. A. & Brossard, D. Believing and sharing misinformation, fact-checks, and accurate information on social media: The role of anxiety during COVID-19. *New Media Soc.* **25**, 141–162. <https://doi.org/10.1177/1461448211011451> (2023).
73. Smith, E. R. & Conroy, F. R. Agent-based modeling: A new approach for theory building in social psychology. *Pers. Soc. Psychol. Rev.* **11**, 87–104. <https://doi.org/10.1177/1088868306294789> (2007).
74. Gausen, A., Luk, W. & Guo, C. Can we stop fake news? Using agent-based modelling to evaluate countermeasures for misinformation on social media. In *Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media (ICWSM)*. <https://doi.org/10.36190/2021.63> (AAAI, Palo Alto, 2021).
75. Pilditch, T. D., Roozenbeek, J., Madsen, J. K. & van der Linden, S. Psychological inoculation can reduce susceptibility to misinformation in large rational agent networks. *R. Soc. Open Sci.* **9**, 211953. <https://doi.org/10.1098/rsos.211953> (2022).
76. Iizuka, R., Toriumi, F., Nishiguchi, M., Takano, M. & Yoshida, M. Impact of correcting misinformation on social disruption. *PLoS ONE* **17**, e0265734. <https://doi.org/10.1371/journal.pone.0265734> (2022).
77. Coscia, M. & Rossi, L. How minimizing conflicts could lead to polarization on social media: An agent-based model investigation. *PLoS ONE* **17**, e0263184. <https://doi.org/10.1371/journal.pone.0263184> (2022).
78. Burbach, L., Halbach, P., Ziefle, M. & Calero Valdez, A. Who shares fake news in online social networks?. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* 234–242. <https://doi.org/10.1145/3320435.3320456> (ACM, New York, 2019).
79. Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9216–9221. <https://doi.org/10.1073/pnas.1804840115> (2018).
80. King, G., Schneer, B. & White, A. How the news media activate public expression and influence national agendas. *Science* **358**, 776–780. <https://doi.org/10.1126/science.aao1100> (2017).
81. Media bias ratings. <https://www.allsides.com/media-bias/ratings>.

82. Goebel, J. T. & Susmann, M. W. On the limits of context-free information theory: novelty and belief consistency as separable predictors of information sharing. *Open Science Framework*. [https://osf.io/wy3hp/?view\\_only=b4f98d8ddf746a1a71df412ee829a6a](https://osf.io/wy3hp/?view_only=b4f98d8ddf746a1a71df412ee829a6a) (2023).
83. 2020 report cards. <https://www.govtrack.us/congress/members/report-cards/2020> (2020).
84. Blankenship, K. L. & Wegener, D. T. Opening the mind to close it: Considering a message in light of important values increases message processing and later resistance to change. *J. Pers. Soc. Psychol.* **94**, 196–213. <https://doi.org/10.1037/0022-3514.94.2.94.196> (2008).

## Acknowledgements

The authors would like to acknowledge the following sources of funding from the National Science Foundation - ITE-2137806 (R. K. G., M. W. S., and S. P.) and SBE-1949037 (S. P.). Any opinions, findings, and conclusions are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## Author contributions

Conceptualization: J. T. G., M. W. S., S. P., H. E., D. T. W. Data curation: J. T. G., M. W. S. Methodology: J. T. G., M. W. S., S. P., H. E., D. T. W. Resources: J. T. G., M. W. S. Investigation: J. T. G., M. W. S. Visualization: J. T. G., M. W. S. Supervision: S. P., H. E., R. K. G., D. T. W. Writing—original draft: J. T. G., M. W. S., S. P., H. E., R. K. G., D. T. W. Writing—review and editing: J. T. G., M. W. S., S. P., H. E., R. K. G., D. T. W.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-56086-2>.

**Correspondence** and requests for materials should be addressed to J.T.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024